Weekly Report

Lu Junhua

2016 年 5 月 30 日

Done

- Gave a lesson to Singapore students.
- Finished code assignment 2,3, and 5 of applied maths in CS.
- Discussed with Prof. Wu. He proposed two themes for my future work on game data: game data mining & anomaly pattern detection in MMORPGs. I read several papers regarding these topics. To be honest, I spent much time on the abovementioned code assignments. Thus, I made brief reflections below, and I will make more delicate surveys on these topics in the next week(s).

To do

- More focus on the game data related papers.
- Finished course assignments (translation and code 4) before June 1st.
- Learn more about javascript

About the two themes

The data mining here in game data mainly refers to predictive analysis. Many indicators (比如用户流失, 付费率等指标) can be predicted via classification or similar techniques. The indicators can be found on game analysis websites, and 祁烨had a brief summary on these long long ago. I will list the common indicators that have been studied next week.

Anomaly detection, in the context of online games, falls into the topic of bots detection and related scenarios. I haven't found out papers related to money laundry or other money-related topics (比如说刷道具拿来卖钱, 中奖拿来卖钱, 这也可能是不算非法或者不对游戏造成太大危害的而不属于游戏作弊; 当然这些也有可能有bots实现因而归入前面一个问题中). In fact, the techniques used in anomaly detection are data mining techniques as well, but usually in a different form compared with standard forms.

A professor (陈宽达) from 台灣中央研究院has made many researches on game data "http://www.iis.sinica.edu.tw/swc/research/games.html". Both two themes are included in his publications. I read several works of his. "On Prophesying Online Gamer Departure" uses SVM to classify players and for potential 流失用户预测. Two bots detection papers use different methods. The first one uses trajectory analysis combined with manifold learning (那么这里城市轨迹数据的一些方法其实也可以用), and the second one analysis the behavior type and frequency to detect. 这两种方法看起来都不错, 但对我们来说还是有问题的. 首先, 我问了网易之前和我们合作的人, 表示并没有轨迹数据; 其次对于第二种, 行为是和游戏上下文非常紧密相关

的, 如果意图采取类似方法的话需要对游戏进一步调研. 从我直观角度来看, 这些方法也不算复杂, 但看看到作者效率评估, 感觉还不错(这和数据量应该也有关系); 此外这些文章多是较早以前的, 我需要在后面找一些时间更近的文章来看看现在的一些分析方向.

Regarding visualization, there is not much work that are quite close to anomaly detection in game data (在我当前认知下). The related work of the paper "TargetVue: Visual Analysis of Anomalous User Behaviors in Online Communication Systems" shows that much work of visualizing anomaly patterns are focused on social networks (除了一般想得到的事件爆发外, 还有水军或者发帖机social bot之类). And for the limited information on players' communication of our data, I think we may not focused on the method related to communication. Also the related work provides many anomaly detection algorithms and a famous survey. Most visualization on these algorithms are naive (有些可能本身这个方法就是这样, 比如通过频次直方图来判断是否有异常). 如果我们能在算法适用的情况下把这些naive的可视化做的更好一些, 或者把这些社交上的anomaly detection 方法转换过来使用, 也许是条路子.

接下来的时间, 我应该要做的一些事情, 从我直觉上想到的: 再找找常见的用于预测的指标, 及对应的方法论的教新的文章; 游戏异常检测, 游戏作弊有哪些分类, 每种分类下的检测应对机制是怎样的. 为了避免过于狭隘, 还要尽可能收集相关方法, 思考其是否可以拿过来运用(比如上面提到的social bots detection的方法能不能用来探索别的类型的bot detection? 因为有一个social bot challenge的比赛, 每年都有). 在之后就是问题的细化和实践. 陈宽达网站上提供了一个魔兽世界的数据集, 后面我也可拿来探索和试验.